AD

TECHNICAL REPORT 7907

GUIDELINES FOR ACUTE TOXICOLOGICAL TESTS

JERRY W. HIGHFILL

US ARMY MEDICAL BIOENGINEERING RESEARCH and DEVELOPMENT LABORATORY
Fort Detrick
Frederick, Md. 21701

NOVEMBER 1979

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION UNLIMITED.

US ARMY MEDICAL RESEARCH AND DEVELOPMENT COMMAND
FORT DETRICK
FREDERICK, MD 21701

NOTICE

Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Disposition

Destroy this report when it is no longer needed. Do not return it to the originator.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
TECHNICAL REPORT 7907

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
GUIDELINES FOR ACUTE TOXICOLOGICAL TESTS

5. TYPE OF REPORT & PERIOD COVERED
Final Report
Oct 1977 — Sep 1978

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
JERRY W. HIGHFILL

8. CONTRACT OR GRANT NUMBER(s)

9. PERFORMING ORGANIZATION NAME AND ADDRESS
US Army Medical Bioengineering Research & Development Laboratory, ATTN: SGRD-UBG, Fort Detrick, Frederick, MD 21701

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61101A
3A161101A91C/00/151

11. CONTROLLING OFFICE NAME AND ADDRESS
US Army Medical Research & Development Command
ATTN: SGRD-AJ
Fort Detrick, Frederick, MD 21701

12. REPORT DATE
November 1979

13. NUMBER OF PAGES
27

14. MONITORING AGENCY NAME & ADDRESS (If different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Acute tests of toxicity
Confidence interval
Guidelines
Logit
Median effect level
Probit
Report requirements

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Drawing from the similarity of acute testing using plants, fishes, and mammalian species, a discussion of experimental designs is presented. Data analysis with two numerical examples is included along with minimum reporting requirements from acute testing.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

SUMMARY

This guide has attempted to provide general fundamental considerations for acute testing in the broad areas of aquatic, plant, and mammalian toxicity testing. It is intended to support the experimental considerations of each field and point out some additional statistical aspects neglected in related guides and often omitted in reports.

Minimum report requirements are listed and described. Graphic displays of short-term toxicological testing summarize results vividly and succinctly. Since many computer programs do not include data displays, discussion and examples of display details are included.

One example indicates methods of adjusting estimates when individuals in the control group also respond. Another example presents tests for the consistency of dose response when combining results from more than one experiment.

Sections included are the selection of dosage levels, randomization, selecting computer programs for processing data, suggestions on the data analysis, and restrictions on the uses of the techniques.

A

PREFACE

From even relatively simple toxicological tests many possibilities exist for combining information and for presenting summary results. As related information is being organized, evaluated, and extracted, it is important that each report be accurate and complete. Having reviewed reports from many institutions performing a wide range of toxicity tests for the US Army Medical Research & Development Command the need for completeness of reports has surfaced. Some shortcomings are conveying only portions of the statistical analysis, improperly using statistical techniques, and missing the necessary adjustments in data analysis when modification of the laboratory tests are required by some preliminary experimental result or a limitation of facility or staff.

The purpose of this document is to provide researchers doing work for the Environmental Protection Research Division of the US Army Medical Bioengineering Research and Development Laboratory (USAMBRDL) some guidance in designing, analyzing, and reporting acute toxicology data. Minimum experimental and reporting requirements are included. This guide may help bridge the gap that many researchers find when working directly with the statistical literature.

A conscious attempt has been made to recognize the intended uses of the results from acute tests, the limitations of mathematical statistics, and the limitations of the experimenter.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

It is probably most helpful to begin by considering how a large population of individuals (mammals, fishes, invertebrates, plants) would respond to a short-term exposure. Different tolerances are expected to be exhibited by different individuals, leading to a distribution of tolerances, the population average, μ and the population variance, σ^2. Often this distribution will be skewed. By making a change in scale, e.g., the logarithm of the value, a more symmetrical distribution may appear. This is advantageous for many reasons but does add a level of complication.

When trying to estimate the tolerance distribution, one recognizes that the cummulative tolerance distribution function is the one actually being estimated by toxicity tests. When a test is performed on 20 animals at a concentration of 23 mg/kg, one observes those animals sensitive to 23 mg/kg or less of the substance.

Upon plotting the proportion affected versus the concentration, it becomes obvious that the plot is not approximated by a straight line. Experience suggests that two transformations be used. One is taking the logarithm of the concentration and the second is a transformation on the proportion affected. A commonly used transformation on the proportion is the probit (coined from probability unit). By finding the normal equivalent deviate of a standard normal distribution (often denoted by z) with area equal to the observed proportion, the observed probit can be found by adding 5 units. Tables also exist for obtaining the probit value directly.[1]

Instead of using the normal distribution, the logistic distribution can be used and a logit results instead of a probit. The logit is a logarithmic function of the number responding and failing to respond. Its use in toxicity is infrequent when compared to the probit.

The typical acute toxicity test will consist of groups of individuals exposed to several concentrations of a substance. From this experiment estimates of μ will be determined. It will be the concentration where one half of the individuals are estimated to respond (EC50). Additionally, fiducial limits, the values where true mean μ is relatively certain to be, will be called confidence limits. EC10 is the estimated concentration where 10 percent of the individuals would be expected to respond.

If death is the response of interest, the estimated concentration where lethality corresponds to 50 percent of the group is denoted by LD50. In aquatic tests concentration replaces dose and LC50 is used. The more general term EC50 (EC = effective concentration) will be used to represent estimates of this nature. Dose and concentration will be used interchangeably because of the general nature of this report.

MINIMUM REQUIREMENTS FOR ACUTE TESTS SUBMITTED TO USAMBRDL

The first and most important requirement is that the experiment be performed in a scientifically acceptable fashion and within the guidelines provided by other documents.[2-4] A random assignment of test organisms to concentration levels and provisions for controlling other environmental differences or time effects such as chemical changes, evaporation, or changes in the sensitivity of the group of individual during the period of dosage, is also required. No animal used in a prior experiment should be used in another acute study.

It is required that a graph showing both the probits of the observed fraction affected and the estimated probit line be presented. Either in a table or in a paragraph, the following information must also be provided: The observed responses and their associated concentrations, the transformation used on the concentration scale, the slope and the intercept of the probit line, the EC50 and its 95% confidence limits in the original concentration units, and the source of the computer routine used to provide the estimates.

In order to avoid extrapolation beyond the concentrations observed, it is required that an estimated EC50 value be given only when a value less than 50% has been observed (0% is acceptable at a tested concentration) and that a value greater than 50% has been observed (100% is acceptable). Experiments leading to only extrapolated values must be repeated. Ideally, dosage levels should be chosen so as to produce little or no response at the lowest dosage level and nearly total response at the highest dosage level.

If the EC10 is needed, as in planning repeated dosing mammalian studies, then an observed response less than 10% is required (0% is acceptable at a tested concentration).

A partial response occurs when, at a given concentration, at least one, but not all, of the individuals respond. An experiment with less than two partial responses should be repeated. This is necessary not only because of mathematical problems in estimating the EC50 and its confidence interval, but because there is less internal confirmation of toxicity in the concentration region of interest. When repeating an experiment that has provided less than two partial responses, different concentrations should be used on the next experiment. See the section on experimental considerations for suggestions.

It is expected that within reasonable sampling variability that an increasing proportion of individuals will respond to an increasing concentration of a compound. When more than two partial responses are observed, an assessment of the goodness of fit of the data to the probit model can be carried out by a chi-square (χ^2) test. When judged to show large inconsistencies of responses to increasing concentration, the experiment should be repeated. More care in specifying subject size or other selection criteria, more control of exposure conditions, different exposure levels, and other experiment plans should be considered.

Sometimes a large χ^2 test statistic indicates a systematic departure. Corrective transformations should be used. If no satisfactory transformation can be found, then the estimate of the EC50 and confidence interval should be given and identified as not having satisfied the goodness of fit test. Finney[2] on page 72 discusses the use of a heterogeneity factor for confidence interval adjustment. If this adjustment is used on the interval, it should be labeled as such.

If animals in the control group respond, that response is called a natural response or, less frequently, a threshold response. Adjustments in the responses of the individuals in the other concentration groups are necessary. In Example 1 of the last section some details are discussed but more than three partial responses will be necessary for goodness of fit test if the adjustment is formally done by the statistical estimation procedure frequently called the optimum pro..dure.

EXPERIMENTAL CONSIDERATIONS

Randomization

The process of randomization is s. central to the running of experiments that a description of assigning .2 individual to 4 treatment groups will be presented in detail. It is assumed that the 12 being used have been obtained from a larger group where unhealthy individuals have already been discarded and that the remaining individuals are relatively homogeneous in size and age.

Suppose a random number table has been chosen and it is decided to use the third page, second column, the right most two digits of that table. The di.its 00 and 13 through 99 will be ignored. The first five digits are 7, 12, 6, 4, and 2. Note that no number is used twice. A number that appears again is ignored. Using two columns, one for each number desired and the other to represent the order the number was obtained, the following is seen:

1	7
2	12
3	6
4	4
5	2
6	
7	
8	
9	
10	
11	
12	

By drawing a slash through each number in the first column, duplicating any number is less likely.

After all the numbers are drawn, the list is as follows:

1	7	
2	12	T_2
3	6	
4	4	
5	2	T_o
6	11	
7	10	
8	5	T_1
9	1	
10	9	
11	3	T_3
12	8	

Choosing four more random numbers from 0 to 3, the sequence found was 2, 0, 1, and 3. These numbers correspond to T_2, the middle concentration; T_0, the control; T_1, the low concentrate; and T_3, the upper concentration, as labelled above. Each of these letters with subscript appear to be right of each group. Finally, as an administrative convenience, each of the numbers in each group is ordered from smallest to largest.

T_o	T_1	T_2	T_3
2	1	6	3
4	5	7	8
11	10	12	9

The first animal selected will be treated with the low concentration, the second animal will be treated as a control, and the third animal will receive the highest concentration. The rest of the sequence should be obvious. For moderate size experiments small cards are a useful organizational aid. For large experiments or when several similar sets are needed, the use of a computer and random number generator is economical. Regardless of the process, the numbers should be carefully checked to see that each number appears only once.

The process just described is called a completely randomized allocation of individuals. The requirement of randomization is usually not convenient, but the protection and the mathematical properties provided by it more than justifies the experimental inconvenience.

Occasionally a stratified random sample may be preferable to a completely randomized allocation of individuals. This would be the case if an investigator needed to systematically control for certain factors such as littermates or body weight.

For example, suppose 12 animals are to be assigned to 4 groups (3 per group) and systematic control for body weight differences is needed. One could then employ a stratified random sampling procedure as follows: Weigh all 12 animals. Take the four heaviest and assign one to each concentration group at random (using a table of random numbers as previously described). Repeat the process for the four middle and four lightest animals. The resulting assignment will be a stratified (by body weight) random sample.

The important role of randomization does not stop once the individuals are treated. Other environmental factors that could influence what will be called treatment effects later must be guarded against. For instance the different treatment groups should not be exposed to varying amounts of light, humidity, temperature, or other relevant factors. The use of randomization requires more vigilance and care for technicians and investigators. Records of the exact randomization procedures used must be kept and different randomization lists should be used for each experiment.

Controls

Control individuals will be used in each test of aquatic organisms and their use is encouraged in other testing such as with mammalian species. The allocation of the individuals as controls will be done at the same time that animals are assigned to treatment regimens. The process will be either a completely random process or a stratified random sampling procedure. A discussion of this process is given in the randomization section.

The individuals in the control group will be treated as closely as possible to the individuals in the treated groups, except for the exposure to the compounds under study. If solvents are needed to deliver the compound, then the controls will be exposed to a solvent level at least as large as the highest solvent level used in a treated group. In addition, if the solvent used is not documented as being used on similar species at similar levels, then an additional control group not receiving the solvent must be used.

Occasionally the control group will respond at a rate higher than zero. An adjustment is appropriate in these cases. If no adjustment is made, then the compound being tested will be estimated to be more toxic than was apparent in the experiment and the chi-square test is likely to indicate model problems. Two approaches are possible. The recommended approach is to use a program that will estimate the natural response rate and make the necessary adjustments in all other estimates. The second approach -- to be used only if the natural response rate is greater than zero but less than 10 percent -- is to adjust the results appropriately. In some cases this will mean that values observed in the lower concentrations will be combined with the value observed in the controls. The new estimate of the response from the controls can then be used to adjust all other responses and a probit analysis can be done on the adjusted values. See Example 1 of the last section for details.

If corrections for natural mortality are necessary, the results should clearly state what decisions were made if low doses are combined with the control and which program was used in the subsequent statistical analysis.

Selection of Dose Levels

Suppose it is known from a small sample range finding study that the highest concentration to be used in an acute aquatic test is 100 mg/l. Range studies with a few individuals at each dose are encouraged as being economical, helpful in planning a more formal acute test, and give supportive information on future findings. It is desirable to have the doses equally spaced on a logarithmic scale (see below). Considering the following sequences, notice that each number is obtained by multiplying its closest left row member by the number in the parenthesis.

(0.9)	100	90	81	72.9	65.6
(0.7)	100	70	49	34.3	24.0
(0.5)	100	50	25	12.5	6.2
(0.3)	100	30	9	2.7	0.8
(0.1)	100	10	1	0.1	0.01

Decisions on the sequence of choice will depend on the information available and the explicit purpose of the experiment. Ideally, the highest concentration would affect all or nearly all of the individuals. Each of the rows of numbers presented above represent equal spacing on logarithmic scales.

The lowest concentration would be determined by the experimental objectives. for example, the level where environmental exposure could reach in spills, just below the expected no effect level, or if the interest is only in the EC50, the point where at least a few individuals are expected to respond.

Finney[1] recommends in experiments intended to estimate EC50 that the number of animals to be tested be divided equally among the doses to be tested. If the location of the EC50 values is actually unknown and is expected to lie anywhere within the specified concentration to be tested, this is a reasonable choice.

If estimates of the EC50 are available and if the purpose of the proposed study is to check the median response rate Finney[1] on page 144 indicates the optimum choice under equal allocation to minimize the variance of the EC50. For instance, 30 animals would be recommended at each of three doses corresponding to EC20, EC50, and EC80. A corresponding experiment with two doses and 18 animals would suggest nine at EC15 and nine at EC85. The designs would be expected to minimize the variance of the EC50 for the number of animals specified. A larger number of individuals in each group would be expected to reduce the variance even more, but another arrangement of doses might do even better than just increasing the number in each group.

The use of measured concentrations during testing is preferred over nominal concentrations. Suppose that 10 individuals were to be dosed with a chemical at 10 mg/m^3 in air and groups of five were to be exposed together. If the measured concentrations were 8.5 mg/m^3 and 9.7 mg/m^3, the results should be calculated with the two concentrations and their individual responses. The extension of this recommendation holds even if each animal was dosed individually.

An exception to using the measured concentration would be if the method measuring the dose was more accurate than the method of sampling the dose. Adjustments for volatility, when needed, should be made when specifying the dose received. These adjustments could be necessary in some acute aquatic bioassays when using highly volatile materials.

Sample Size

Minimum sample size is often specified in guidelines written for specific types of tests of toxicity. Several potentially conflicting goals exist when one plans an acute test for toxicity. One is to establish a dose-response relationship over several concentrations. Another is to get a usable estimate of the EC50, and often there is interest in the concentration where few individuals are affected. Interest in levels where a few respond could be with aquatic organisms directly exposed to a contaminant or in planning from a single acute exposure for multiple exposures.

In considering the sample size necessary to estimate the EC50 within a certain distance of the true population value, some important factors need to be considered. One is an estimate of the variances and another is the direct relationship of concentration values being tested and the responses to be observed. It is obvious that a very large number of individuals tested at poorly chosen concentrations will provide an undesirable estimate of the EC50 or no possible estimate. A less obvious consideration is that the responses at different concentration receive different mathematical weights. This directly effects the estimation of sample size. DeArmon and Lincoln suggested that an average mathematical weight of 0.40 be used.[5] The suggested value came from a large number of experiments with pathogenic infectivity.

If preliminary information exists so that estimation of the slope of the probit line can be obtained, the total number of individuals required can be estimated. To be 95% certain that the estimate is within D units of the population value given an estimate of the slope as B, the total number needed is $19/(2\ B^2\ D^2)$ where $2D = L$ in DeArmon and Lincoln's paper. This total number could then be divided among each group to be exposed as various concentrations. If serious interest existed in either extreme of the concentration range, additional individuals could be added at the point of interests or additional future experiments could be planned at these points.

ANALYSIS OF RESULTS

Analysis of the Data

Some acceptable methods of analysis of acute dose response data are the logit, probit, and a nonparametric method such as Spearman-Karber.[1] The probit method is the most readily available at computer centers. Suggested options of computer programs are discussed in the Selecting Computer Programs Section.

Briefly, the statistical model has been based on the following assumptions: (1) That some transformation of the concentration (usually logarithmic) allows the function of observed responses to be approximated by the cummulative normal distribution and (2) individuals receiving the dose respond independently according to the binomial distribution where the parameter relates to normal distribution and concentration level. Using these assumptions, the statistical method of maximum likelihood is used to find estimates of the slope and intercept on transformed scales. Since the solution to this problem is not explicit, the actual process is a bit messy. Fortunately, computer programs do most of the drudgery.

The solution process is fairly straightforward if at least three partial responses are observed, and each tends to increase with higher concentrations.

It is expected that the dose level will be transformed logarithmically either before the analysis or by the computer program. It is also expected that the results will be transformed back into their original units. Two exceptions to using logarithmic transformations are when the dose levels have been chosen very close together and when curvilinear results are observed upon graphing the experiment. In the case of curvilinearity other transformations will be necessary to satisfy the assumption of linearity. When the doses are very close, say within an order of magnitude, no transformation of concentration may be necessary.

<u>Selecting Computer Programs</u>

The following features are desirable when looking for a computer program to do either probit or logit analysis of the results from acute testing:

(1) If nonlinear responses are observed, transformations other than logarithmic may be necessary. Special care must be taken in these situations when expressing the confidence interval and EC50 in their original units. Programs that handle transformations of the concentrations and convert meaningful statistics to their original concentration units are helpful.

(2) Assuming a weighted regression routine, provide estimates of the intercept and slope of the fitted probit or logit curve.

(3) Provide a chi-square goodness of fit test for heterogenity of responses.

(4) Handle situations where nonpartial responses occur. This may not be obvious when checking written descriptions of programs but needs to be verified.

(5) Calculate estimates when a natural response is observed in the control group.

(6) Plot the responses observed and the fitted line from the computer routine.

It is strongly recommended that contractors do not develop programs to do probit analysis for data gathered for USAMBRDL. A program run on SAS through the NIH computer center costs less than one dollar.[6] The BMD program should not run much higher.[7] Any program development would be much more expensive than the rental or usage fee of prepared programs.

With today's access to computers, there is no reason to accept a graphical approximation such as the Litchfield-Wilcoxon.[1]

Combining Duplicates

Often in aquatic assays duplicate tanks will be run to increase the sample size needed. If in the flow through apparatus adequate mixing and splitting takes place, the results can be pooled and the analysis performed as though one large group had been done at each concentration. The same would be true in a static test if the two tanks being used were pumped from a larger well mixed source.

For plants being exposed to aerosols several precautions should be taken before further considerations of combining results. These are that the concentrations being maintained were in fact the same, that seasonal differences have not occurred, that the plants came from the same group and were essentially the same age, and that the times of exposure each day were about the same.

Similar warnings should be used when considering combining information from mammalian studies. These might include the use of different technicians, shipments of animals, age or weight of animals, times of exposure and other factors dependent upon mode of exposure.

In every case the number exposed, the number responding, and the concentration used must be provided by chronology or other satisfactory labeling before presenting the combined information.

If more than one compound is to be tested for comparative purposes, other more important experimental considerations should not be overlooked in the experimental design. Since this type of testing is beyond the scope of this document, the use of professional statistical collaboration is suggested.

As concern mounts over whether duplicates or even replicates are responding differently, tests of significance between batches are recommended. Finney on page 176 shows how this test can be performed. Also see Example 2.

Related Problems and Restrictions to the Methods of Estimating Median Effect Levels

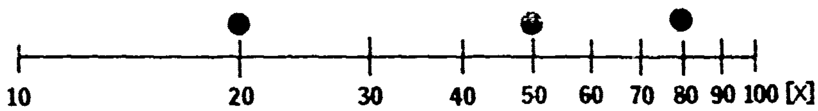
As was stated earlier, the number of individuals responding divided by the number exposed is converted to fractions and transformed into probits and a weighted regression is calculated. It is assumed that the number responding will follow the binomial distribution. An important assumed property when using the binomial distribution with a group of individuals being exposed is independence (usually assured by randomization) of the respondent. In the case of acute studies on plants, one might be tempted to use the response as the number of leaves responding out of the total number of leaves exposed. Important drawbacks to using the methods described in the analysis of data section are that individual leaves cannot ordinarily be assigned at random to the treatment groups and that the response of individual leaves on a plant is likely to be very dependent upon the sensitivity of other leaves growing on the same plant. Two important problems evolve: What weights to use in the analysis and the effect of dependent responses. In the particular case it is recommended that the investigator check to see if the number of exposed leaves in each treatment group is approximately the same. If it is, then a transformation such as the probit, arc sine, or odds ratio of the responses to linearize the response to increasing dose is recommended. It can be followed with a standard unweighted regression analysis. The analysis of dependent responses needs more research.

Another problem that has been observed is the mistake of thinking that 20% means that 100 individuals were used and that 20 responded. The following responses to number exposed gives 20%: 1/5, 2/10, 200/1000. Only the actual numbers making up the ratio should be used. When each number has to be estimated as in a microbiological study, then the method of Wadley in Finney,[1] page 202, should be followed.

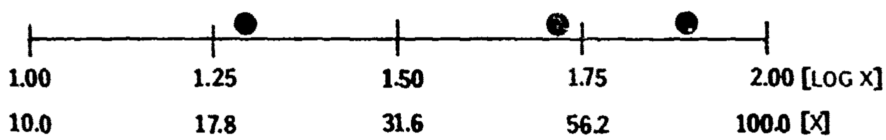
In situations where the effect being measured occurs quickly, the sequential method called the up-and-down design should be considered.[1,8] The process is simple. If a subject responds at a concentration, the next subject is given a lower concentration. If the subject does not respond, the next subject is given a higher concentration than just given. This method is designed to estimate an EC50 value but gives a less satisfactory estimate of the variance, σ^2. This method allows for a more efficient use of experimental subjects. This method would not be recommended if the EC20 or a lower value were required. Once the data are collected, they can be analyzed by the probit or logit method.

EXAMPLES OF PROBIT ANALYSIS

When plotting observed values and their corresponding concentrations, one has several choices. The first is to make all transformations and then plot on ordinary graph paper. Another is finding the proper graph paper and plotting the numbers observed. The following example shows the process for plotting concentrations on logarithmic paper and plotting the logarithms on standard paper. Suppose we wish to plot 20, 50, 80 (upper); Log 20 = 1.301; log 50 = 1.699; and log 80 = 1.903 (lower).



Logarithmic Scale



Ordinary scale plotting logarithm of values

Graph paper exists with normal probability on the vertical axis and logarithm, base 10, on the other. One can use that paper by the following procedure. Mark off the horizontal axis in powers of 10. Assuming that the calculated probit from the program is $Y = \hat{\alpha}+\hat{\beta}\log X$ where $\hat{\alpha}$ is the intercept, $\hat{\beta}$ is the slope, and a base 10 logarithm transformation was used on the data. Let $(6-\hat{\alpha})/\hat{\beta} = X1$ and $(4-\hat{\alpha})/\hat{\beta} = X2$, find 10^{X1} and 10^{X2}. For notation purposes let X3 equal the value of the first power of 10 and X4 equal the second. Plot the points (X3, 0.84) and (X4, 0.16). Connect these two points with a straight edge and extend in both directions. The line just formed is the expected probit line on normal probability paper with a logarithmic horizontal axis.

By next plotting the observed percentage responses from each concentration group, the required graph is formed. Example 1 shows a case where a natural response rate was observed in the controls. In that case a corrected percentage response would be used instead of the observed percentage response. Using the equation $Y = 3.37 + 4.38 \log X$, one finds $(6-3.37)/4.38 = 0.60$, so $X3 = 3.99$ and $(43.37)/4.38 = 0.14$, so $X4 = 1.39$. The examples included in this section have used the transformed values on both axes and have been plotted as on ordinary (Cartesian) graph paper.

In this section, an estimate of the EC50 is given with upper and lower bounds. The bounds in each case are 95% confidence intervals.

Example 1, The Effects of a Natural Response

Table 1 shows the results from the exposure of fathead minnow (Pimephales promedas) eggs to nitroglycerin at various concentrations for 144 hours.[9]

TABLE 1. THE EFFECT OF NITROGLYCERIN ON NUMBER OF FATHEAD MINNOW EGGS HATCHED AFTER 144 HOURS

Set	X Conc.	No.	No. Not Hatched	Fraction Responding	$\frac{\text{Fraction}-0.084}{1-0.084}$ [a]	Est. Probit	Probit of observed
	Controls	30	3	0.100			
1	0.10	30	1	0.033	-0.056		3.16
2	0.16	30	2	0.067	-0.019		3.50
3	0.24	30	2	0.067	-0.019		3.50
4	0.37	30	3	0.100	0.017	2.88	3.72
5	0.56	30	3	0.100	0.017	2.88	3.72
6	0.87	30	6	0.200	0.127	3.86	4.16
7	1.20	30	6	0.200	0.127	3.86	4.16
8	1.80	30	7	0.233	0.163	4.02	4.27
9	2.80	30	21	0.700	0.672	5.44	5.52
10	4.20	30	27	0.900	0.891	6.23	6.28
11	6.50	30	29	0.967	0.964	6.80	6.84
12	10.00	30	30	1.000			

$\hat{Y} = 3.37 + 4.38 \log X$

a. Fraction formed is the adjusted value from an estimated natural response rate of 0.084.

Ignoring responses from the controls and running sets 1 through 12, Table 1, the SAS probit program with logarithm of concentration, EC50 = 1.76 with bounds of 1.19 and 2.78. The χ^2 test statistic equals 34.4 with 10 d.f. and α <0.001. This means that a modeling problem has been found with a very small probability of error (<0.001). Inspection of the control data suggests that an adjustment using an estimated natural response rate is required.

Programs exist for optimizing the estimated natural response rate. When this was done using the SAS probit program the estimated natural response rate, C, was found to be 0.084. This is reasonable when one observed 0.10 fractional response in the controls and less at the lower concentration. Other results from the program are: Intercept = 3.37, slope = 4.38, χ^2 = 6.69 with 9 d.f., α = 0.67, EC50 = 2.36 with bound 1.84 and 2.77. If one plots the probits corresponding to observed fractions, it is obvious that predicted line does not fit the data. By considering the fractional responses adjusted for natural response rate, (fraction - 0.084)/(10.084), and their corresponding probits, a reasonable fit is observed. See Table 1 and Figure 1 for details. The details for constructing Figure 1 are given in Example 2.

If a program is not available for optimizing the natural response rate, an adjustment from control values and the lower concentration values is possible. A restriction is that the estimated natural response rate be greater than 0 but less than 10%. If past information suggests that 10% of the control eggs of a fathead minnow do not hatch, under the conditions that the present test was run, then 10% is a reasonable adjustment value to be used (note in Table 1, 0.10 would replace 0.084 in column 6). If the control rate is usually lower, or if the control response appears to be too high, the lower concentration responses can be combined with the control responses to get a more realistic estimate of natural response rate. The point where the combining stops is somewhat arbitrary. In this particular example 8/120 (0.067) failed to hatch at concentrations less than 0.37 mg/l. Making the correction (observed fractional response - 0.067)/(10.067) and running sets 4 through 12 of Table 1, the following results were obtained: Intercept = 3.54, slope = 4.08, χ^2 = 6.87 with 7 d.f., α = 0.44, EC50 = 2.27 with limits of 1.90 and 2.64. A plot of the data showed very little difference in these estimates and the ones obtained under the optimum procedure using data sets 1 through 12.

From this set of data, it is not recommended that the analysis be done on sets 4 through 12 of Table 1 with the optimum procedure. The resulting estimate of C is 0.13 which is unrealistic when compared with the controls and the first three concentrations. C = 0.17 occurs if sets 6 through 12 of Table 1 are used with the optimum procedure. This is even worse.

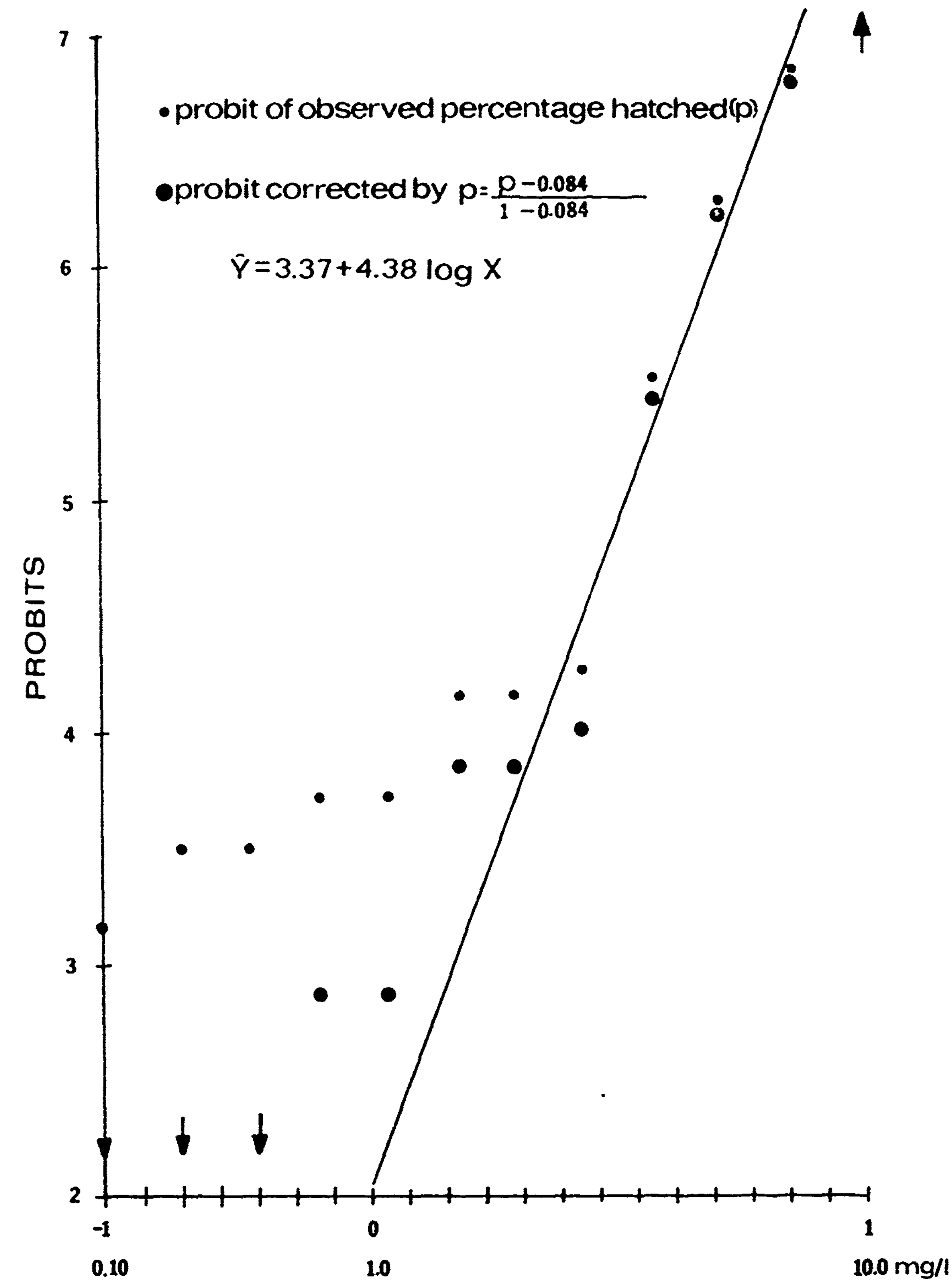


Figure 1. Hatchability of fathead minnow, Pimephales promedas, eggs exposed to nitroglycerin for 144 hours.

It should be pointed out that when a natural response rate is suggested by the data and no correction is made that a lower EC50 is given. This means that the compound is being reported as being more toxic than it should be. When coupled with the economics of environmental containment or treatment, the appropriate handling of information in the control groups can be very important.

Example 2, Combining Replicates and Transformations

When toxicity tests are repeated, differences in responses of different individuals at different times and differences in exposures from one time to another can cause an increased variability.

The following results are from two different days of applying aerosolized acetic acid to tobacco plants (*Nicatiana tobacum*, Speight G28).[10] Only four chambers were available and nine plants per concentration were used on 1 day and 10 plants per concentration on the other day.

The following information was observed:

Concentration (X mg/m^3)	# Plants	# Injured
14	9	2
16	9	4
18	9	5
20	9	7
12	10	0
14	10	2
16	10	5
18	10	6

Although changing concentration is not necessarily a recommended experimental procedure, the data offers an informative example. Coupled with the example on page 176 of Finney,[1] an adequate basis for the combining of replicate experiments should exist.

When the SAS probit program was run using log concentration the following information was entered:

Concentration (X mg/m^3)	# Plants	# Effected
12	10	0
14	19	4
16	19	9
18	19	11
20	9	7

Using the resulting formula, $Y = -8.505 + 11.014 \log X$, for 12 mg/m^3 the predicted value is $-8.505 + 11.014 \log 12 = 3.381$. For the χ^2 test, $\alpha = 0.75$ indicating no departure from linearity. A probit value of 3.381 corresponds to a proportion of 0.053. Following the above guidance for each concentration, columns 5 and 6 of Table 2 were formed. If a probit table is not available, one can use the standard normal distribution with value (3.381 - 5.0) -1.619. The area, or equivalently the probability, corresponding to a -1.62 is 0.053.

TABLE 2. REPLICATES FROM AEROSOLIZED ACETIC ACID ON TOBACCO PLANTS USING LOGARITHM OF CONCENTRATION

X (mg/l)	No. of Plants (n)	Observed Response (r)	Y	$\hat{Y}$	$\hat{P}$	$\frac{(r-n\hat{P})^2}{n\hat{P}(1-\hat{P})}$	$\frac{(r-n\hat{P})^2}{n\hat{P}(1-\hat{P})}$
12	10	0		3.381	0.053	0.560	0.560
14	9	2	4.234	4.118	0.189	0.065	
	10	2	4.158	4.118	0.189	0.008	
	19	4		4.118	0.189		0.057
16	9	4	4.859	4.757	0.404	0.061	
	10	5	5.000	4.757	0.404	0.383	
	19	9		4.757	0.404		0.383
18	9	5	5.138	5.320	0.626	0.191	
	10	6	5.253	5.320	0.626	0.029	
	19	11		5.320	0.626		0.180
20	9	7	5.766	5.824	0.795	0.016	0.016
	$\hat{Y} = -8.505 + 11.074 \log X$					$X_6^2 = 1.313$	$X_3^2 = 1.196$

Following the guidance in column 7, the value $\chi^2 = 1.313$ corresponds to what will be called total residual with 6 d.f. In column 8, the $\chi^2 = 1.196$ corresponds to a measure of deviations from linearity. Batches within doses is found by subtraction. Table 3 summarizes the test for determining increased variability.

TABLE 3. SUMMARY OF LACK OF FIT USING LOGARITHM OF CONCENTRATION FOR ACETIC ACID

Source	Degrees of Freedom	Sum of Squares	Mean Squares
Deviation from linearity	3	1.196	1.196/3 = 0.40
Batches within doses	3	0.117	0.117/3 = 0.04
Total Residual	6	1.313	

The $\chi^2 = 1.196$ with 3 degrees of freedom does not indicate a deviation from linearity. The $\chi^2 = 0.117$ with 3 degrees of freedom shows little variability from repeated observations at the same concentrations. Neither of these values indicates any problems from combining these two sets of information.

As an example of using a transformation of concentration other than logarithmic this same set of data was analyzed as above except the inverse of the concentration was used.

The results from running 1/conc. gave $Y = 9.767 - 79.7/\text{conc.}$, where conc. is concentration in mg/l. Following the guidance given in Table 2, Table 4 was formed using the recalculated coefficients. Table 5 summarizes the information in Table 4 into the appropriate test.

Figure 2 shows the probit equation and the results on scale 1/X. In the transformed scale, EC50 = 59.83 with bounds of 54.53 and 63.92. Dividing each of these quantities by 1000 and then taking the inverse, the estimated EC50 is 16.7 with bounds of 15.6 and 18.3. Using the logarithmic transformation the estimated EC50 is 16.8 with bounds of 15.7 and 18.4.

TABLE 4. REPLICATES FROM AEROSOLIZED ACETIC ACID ON TOBACCO PLANTS USING THE INVERSE OF CONCENTRATION

X	1000/X	n	r	Y	$\hat{Y}$	$\hat{P}$	$\frac{(r-n\hat{P})^2}{n\hat{P}(1-\hat{P})}$	
12	83	10	0		3.125	0.030	0.309	0.309
14	71	9	2	4.23	4.074	0.177	0.126	
	71	10	2	4.16	4.074	0.177	0.036	
		19	4		4.074	0.177		0.147
16	62	9	4	4.86	4.786	0.415	0.032	
	62	10	5	5.00	4.786	0.415	0.298	
		19	9		4.786	0.415		0.269
18	56	9	5	5.14	5.339	0.633	0.232	
	56	10	6	5.25	5.339	0.633	0.047	
		19	11		5.339	0.633		0.239
20	50	9	7	5.77	5.782	0.783	0.001	0.001
	$\hat{Y} = 9.767 - 79.7/X$						$\chi^2_6 = 1.081$	$\chi^2_3 = 0.965$

TABLE 5. SUMMARY OF LACK OF FIT USING THE INVERSE TRANSFORMATION OF CONCENTRATION FOR ACETIC ACID

Source	Degrees of Freedom	Sum of Squares	Mean Squares
Deviation from linearity	3	0.965	0.322
Batches within doses	3	0.116	0.039
Total Residual	6	1.081	

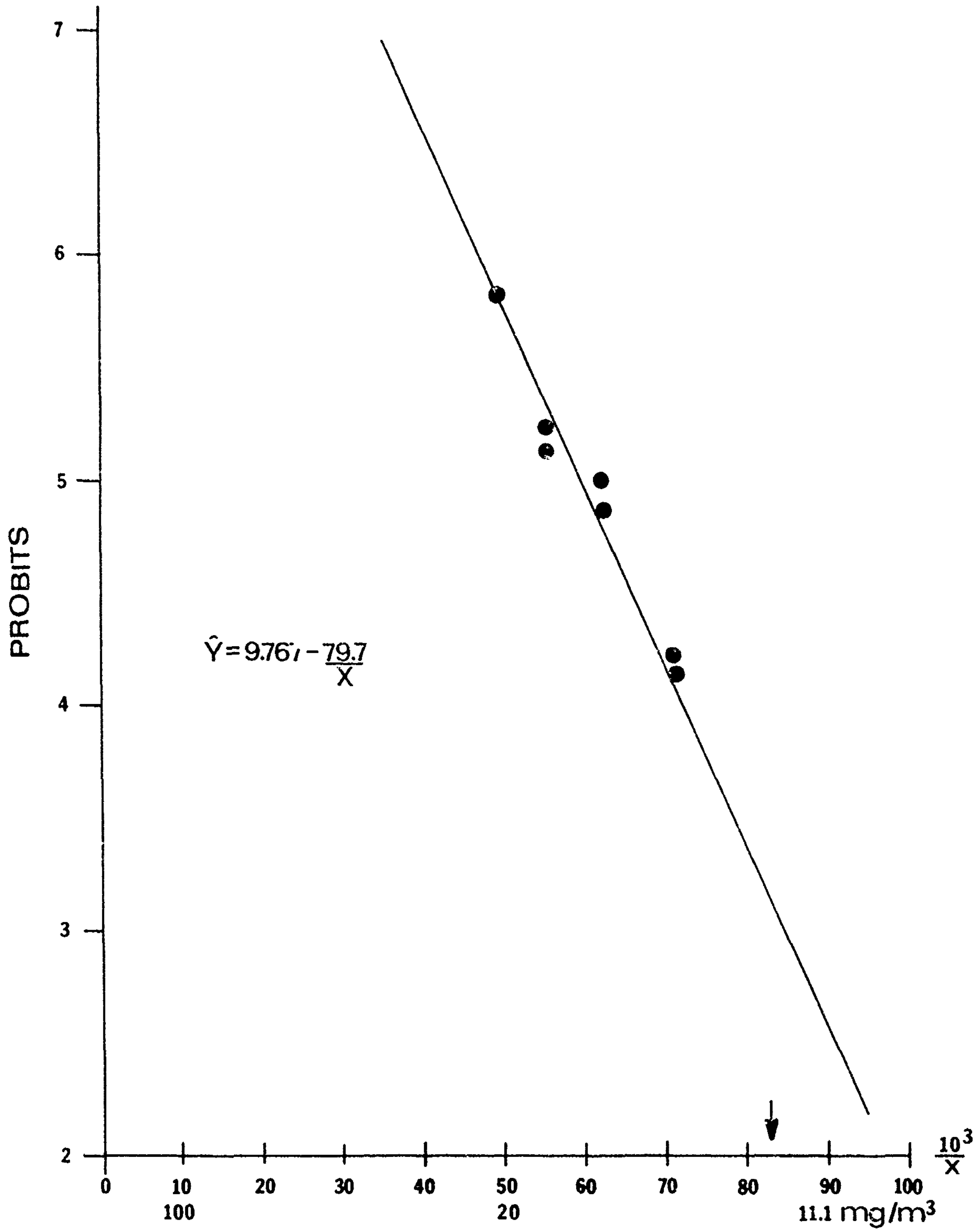


Figure 2. Effects on tobacco plants, <u>Nicotiana tobacum</u>, Speight G28 after 2 hour fumigation with acetic acid.

LITERATURE CITED

1. Finney, D.J., Probit Analysis, Cambridge University Press, 3rd edition (1977).

2. "Guide for the Care and Use of Laboratory Animals," DHEY Publication No. (NIH) 78-23, US Department of HEW, Public Health Service, National Institutes of Health.

3. Sontag, James M., N.P. Page and U. Saffiotti, Guidelines for Carcinogen Biosassay in Small Rodents," National Cancer Institute, Bethesda, MD, Technical Report Series No. 1, Feb 1976 (NCI-CG-TR-1).

4. "Methods for Acute Toxicity Tests with Fish, Macroinvertebrates and Amphibians," EPA-660/3-75-009, National Environmental Research Center, Office of Research and Development, US Environmental Protection Agency, Corvallis, OR (April 1975).

5. DeArmon, I.A., Jr. and R.E. Lincoln, "Number of Animals Required in the Bio-assay of Pathogens," J. Bacteriol., 78(5):651-657 (1959).

6. Barr, Anthony J., J.H. Goodnight, J.P. Sall and J.T. Helwig, A User's Guide to SAS 76, SAS Institute, Inc., Raleigh, NC.

7. Dixon, W.J., Biomedical Computer Programs, University of California Press, Los Angeles, CA (1976).

8. Natrella, M.G., Experimental Statistics, National Bureau of Standards Handbook 91, p. 10-23 (Aug 1963).

9. Bentley, R.E., J.W. Dean, S.J. Ells, G.A. LeBlanc, S. Sauter and B.H. Sleight, III, "Appendix Data for the Laboratory Evaluation of Nitroglycin to Aquatic Organisms," Contract No. DAMD 17-73-C-4101, US Army Medical Research and Development Command, Fort Detrick, Frederick, MD.

10. Thompson, C.R. and G. Kats, "Phytotoxic Hazard of Air Pollution Associated with Munition Production," Contract No. DAMD 17-77-C-7015, US Army Medical Research and Development Command, Fort Detrick, Frederick, MD.

GLOSSARY

Acute toxicology test is where biological organisms are subjected to potential stress or stimulation by materials at one or more levels. The period of exposure can be very brief or as long as a few days.

Control group is individuals similar to those being subjected to a material used in the toxicological test. They are handled in a similar manner to those receiving exposures, but they do not receive the substance under study. In some situations the controls will be exposed to a chemical vehicle used to deliver the substance under study.

EC50 is the estimated concentration where 50% of the individuals would respond. It is used in toxicological tests to estimate the mean of the tolerance distribution. It is used in this paper as a term where E represents effective, lethal, or generally a response and C represents a concentration of exposure or dose delivered.

Goodness of fit in the context of this paper means a statistical measure of departure from the anticipated linearity of responses with increasing exposure at certain ranges of concentration. The term could have been a test of heterogeneity.

Natural response rate corresponds to the proportion of individuals reacting in the control group.

Logit is the natural logarithm of the proportion responding minus the natural logarithm of the proportion not responding at a particular concentration. The estimated logit from a group is modified for results at concentrations where non-partial responses are observed.

Optimum procedure is a statistical estimation technique that adjusts the probit values, and thus the probit equation for situations where a natural response rate is observed in the control group.

Partial response is where at least one, but not all, of the individuals receiving a like stimulation react.

Probit is a transformation of the percentage of responses and is found by taking the area under a standard normal distribution corresponding to the proportion observed. For each area there corresponds a normal equivalent deviate (the numbered standard deviations from the mean), to which five units are arbitrarily added.

Random assignment means that possible assignments of individuals to groups are equally probable.

Tolerance corresponds to a point at which an individual responds to an exposure level. The behavior of a large group, the population, to various concentrations corresponds to the tolerance distribution.

DISTRIBUTION LIST
Project No. 3A161101A91C/00/151

No. of Copies	
5	US Army Medical Research and Development Command ATTN: SGRD-AJ Fort Detrick Frederick, MD 21701
12	Defense Documentation Center ATTN: DDC-PCA Alexandria, VA 22314
1	Academy of Health Sciences, US Army ATTN: AHS-COM Fort Sam Houston, TX 78234
4	USAMBRDL Technical Library